The Montana Department of Environmental Quality Metals Assessment Method

Final

June 2011

Prepared by:
Jonathan Drygas
Water Quality Planning Bureau, Monitoring and Assessment Section
Montana Department of Environmental Quality
1520 E. Sixth Avenue
P.O. Box 200901
Helena, MT 59620-0901

WQPBMASTR-03

Suggested citation: Drygas, Jonathan. 2011. The Montana Department of Environmental Quality Metals Assessment Method. Helena, MT: Montana Dept. of Environmental Quality.

TABLE OF CONTENTS

Acronyms ii
1.0 Introduction 1
2.0 Sampling Frame: 1
2.2 Sample Frame, Population, and Sampling Units 1
2.3 Sampling independence 3
3.0 Assessment methodology 3
3.1 Aquatic Life/Fish Beneficial Use 4
3.2 Drinking Water Beneficial Use (Human Health Standard) 6
3.3 Non-detects and J Flagged Samples 6
4.0 References 7

LIST OF TABLES

Table 3-1. Recommended minimal metals sampling suite. 4

LIST OF FIGURES

Figure 2-1. Four different stream reaches (shown by different colors), each representing 1 sampling frame (ADB stream segment). 3
Figure 3-1. Shows three scenarios involving J flagged data and the Water Quality Standard. 7

ACRONYMS

Acronym	Definition
ADB	Assessment database
DEQ	Department of Environmental Quality (Montana)
EPA	Environmental Protection Agency (US)
MDL	Method Detection Limit
RL	Reporting Limit
SOP	Standard Operating Procedures
TMDL	Total Maximum Daily Load
WQS	Water Quality Standards

1.0 INTRODUCTION

The purpose of this document is to provide an assessment method and monitoring framework to determine compliance with the metals numeric standards as defined in the Circular DEQ-7 (Montana Department of Environmental Quality, 2010).

2.0 SAMPLING FRAME:

For lakes, the sampling frame is applicable to lakes greater than or equal to 1 acre. Lake sampling is conducted at a single location (mid-depth lake). For rivers, the sampling frame is similar to the one for wadeable streams (see below).

The following sampling frame section follows similar guidelines as Appendix A of the "Assessment Methodology for Determining Wadeable Stream Impairment Due to Excess Nitrogen and Phosphorus Levels"(Suplee and Sada de Suplee, 2011). It is important to note that an assessment reach for nutrients may not always be the same for metals but the sampling frame concept is the same.

2.2 SAMPLE FRAME, POPULATION, AND SAMPLING UNITS

All studies involving statistical evaluations of data require that a sample frame, population, and sampling unit be defined. Streams are particularly poor entities for establishing these parameters because streams are an interconnected network rather than discreet entities. Nevertheless, streams are the entities to be sampled so some effort must be made to segregate them into definable units. For the purposes of determining compliance with numeric metals criteria, we define the following:

- **Sample Frame**: *A wadeable[1] stream segment listed in the Assessment Data Base (ADB)* ***OR*** *a sub-segment of an ADB stream segment. These segments are referred to here as an "assessment reach".*

- **Population**: *All the water flowing through the assessment reach during the time period when the numeric metals criteria apply, and the surface area of the stream bottom over which the water flows.*

- **Sampling Unit**: *A sample collected from the assessment reach that is largely independent of other samples collected within the assessment reach and collected during the time when the numeric metals criteria apply.*

Assumptions: Each assessment reach (ADB segment or sub-segment) will be made up of a series of sampling sites, or a series of very short study reaches that are essentially sites (**Figure 2-1**). The minimum number of sites on an assessment reach is provided in DEQ SOPs (Montana Department of Environmental Quality, 2005). **Figure 2-1** illustrates the variety of ADB segments that may be found; segment lengths can vary tremendously. For purposes of determining compliance with numeric metals standards using statistical methods, it is usually assumed that (1) pollution sources are evenly dispersed

[1] Wadeable streams are perennial and intermittent streams in which large portions of the channel are wadeable during baseflow conditions. For the list of waterbody segments not considered wadeable (i.e., the large rivers), see (Flynn and Suplee, 2010).

along the reach, (2) sampling sites are randomly located along the reach, and (3) each sample is independent of the others. Spatial and temporal independence guidelines for sites are addressed in **Section 2.3** below.

It some cases, ADB segments may have pollution problems (hotspots) concentrated only in a particular part of the stream, say, the last 5 stream miles. In such cases, it may not make sense to view the original ADB segment as the best possible sampling frame. That is, it would be better to further stratify the sample frame and, thus, the population of interest. This will prevent distortion of results caused by mixing together, for common analysis, data from the relatively un-impacted sub-segment with data from the impacted sub-segment. For example, in **Figure 2-1** it might be prudent to consider the sub-segment upstream of the Star Mine as a sampling frame apart from the sub-segment below the mine. Stratification is common in studies employing purely random sampling, where it is referred to as stratified random sampling (Cochran, 1977). Stratification allows maximal precision of estimates for minimal sampling effort (Norris, et al., 1992). The assessor carrying out the analysis on an ADB segment will have to judge if further stratification is warranted. If it is warranted, then sampling requirements, described above and further detailed below, would apply to *each* of the new sub-segments (aka assessment reaches), individually.

Precautionary Considerations: Pollution sources are rarely evenly-dispersed along stream segments, violating assumption 1 above. And purely random sampling is usually not practical due to stream access issues, etc. Targeting only the known or potential pollution "hotspots" — even within an assessment reach that has been broken out from a larger ADB segment — could over represent the hotspots and distort the statistical tests. Sampling and analysis plans (SAPs) should proceed with goal-oriented sampling (Buck, et al., 2000) that works towards striking an equitable balance between the number of hotspot sites and the number of un- or minimally impacted sites *within the defined assessment reach*. That is, the aggregate of collected samples should be representative(Environmental Research Laboratory-Duluth, 2002) of the assessment reach as a whole. Advanced knowledge and expertise of the field will be needed to accomplish this (Norris, et al., 1992), and modifications to the assessment reach boundaries can be made on-the-fly during field work, if deemed necessary. It is possible to sub-segment a stream reach to the point where, for a particular assessment reach, there really is little left but hotspots; if this is the case, and the assessor believes it is appropriate, then the hotspots *are* representative of the assessment reach. As a general rule, it is better to lump than split to avoid unnecessary sampling and administrative work. The requirement to create reasonably uniform assessment reaches is inherently in conflict with the need to "lump" for the purpose of keeping assessments as simple as possible. Judgment is needed to balance these two opposed factors and come up with an optimal sampling strategy.

Although this quasi-systematic approach is not a substitute for truly random sampling it will, if carried out properly, achieve good sample interspersion and representativeness. For further discussion of randomization *vs.* interspersion approaches, see page 196 of (Hurlbert, 1984).

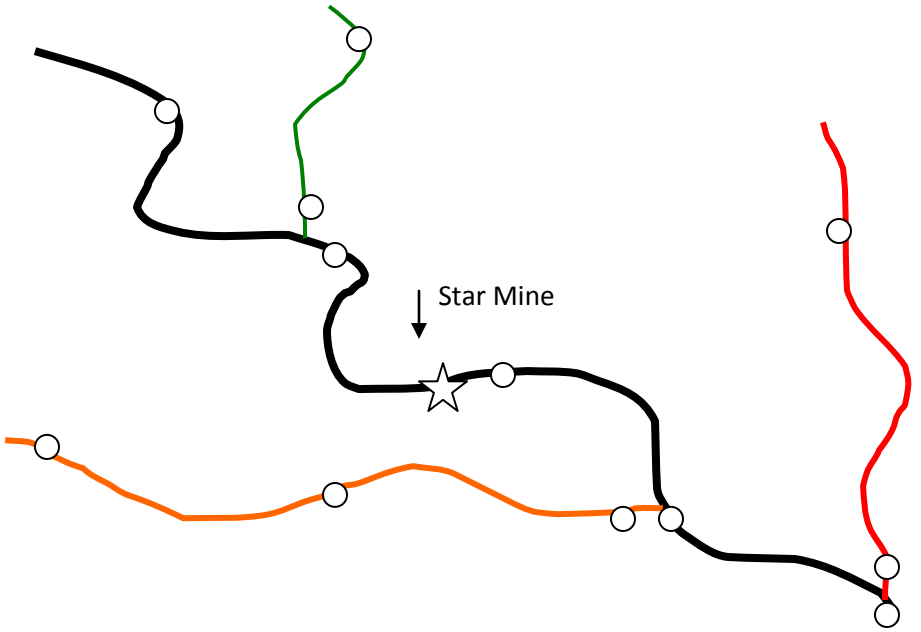


Figure 2-1. Four different stream reaches (shown by different colors), each representing 1 sampling frame (ADB stream segment).
Example sampling sites (hollow dots) are shown along each segment.

2.3 SAMPLING INDEPENDENCE

Sampling independence includes spatial and temporal components. During high flow, the temporal component will be evaluated on a case by case basis, whereas during summer, while streams are at base flow, there should be a 30-day period between sampling events. This is to ensure independence of the samples with an unbiased data set. The spatial component is set at 1 mile between sampling sites. If sources for metals (e.g. tailings piles, abandon or active mines, etc.) are identified, bracketing the sample sites in relation to the source should occur, and in some cases, the 1 mile spatial component might not apply.

3.0 ASSESSMENT METHODOLOGY

When assessing the metals data, both the aquatic life/fish and drinking water beneficial uses should be evaluated. Numeric standards to protect aquatic life and human health are different and therefore the methods on how they are applied differ (Montana Department of Environmental Quality, 2010). In general, some standard exceedances are allowed to assess the aquatic life/fish beneficial use with the exception of silver, which is interpreted as a "not to exceed" value. Just one exceedance is allowed for the drinking water beneficial use (human health standard). Thus, a waterbody can be listed for drinking water but not for aquatic life/fish. In addition, it's important to note that some of the aquatic life standards are dependent on hardness and adjust with changes to the hardness. The minimum sample size is 8 samples within the same assessment reach. Less number of samples might be considered (see specifics for each beneficial use attainment decision). In general, data from the last **ten** years should be only considered when making attainment decisions for both beneficial uses (aquatic life/fish and

drinking water). Data that is older than ten years should only be used as a historical reference and may be looked at for Total Maximum Daily Load (TMDL) development. Samples can be taken any time of the year. Ideally 33% of the data set should be from samples collected during high flow conditions while the remaining should be collected during base flow conditions. A sampling design ratio of one high flow sample per every two base flow sampling events would easily achieve this goal. At a minimum, for non targeted sampling designs, a metals sampling suite should include the Total Recoverable Metals (**Table 3-1**) and dissolved aluminum**.** The rationale for using this subset of metals is based on the most commonly 303(d) listed metals in Montana and their metals associations. Other dissolved and sediment metal sample fractions are not required for making attainment decisions but are valuable in determining potential sources. Thus, specific project needs will determine if these sample collections are necessary or not.

Table 3-1. Recommended minimal metals sampling suite.

Water Sample - Total Recoverable Metals		
Metal	**Method**	**Req. Report Limit (ug/L)**
Arsenic	EPA 200.8	3
Cadmium	EPA 200.8	0.08
Chromium	EPA 200.8	1
Copper	EPA 200.8	1
Iron	EPA 200.7	50
Lead	EPA 200.8	0.5
Selenium	EPA 200.8	1
Silver	EPA 200.8	0.5
Zinc	EPA 200.7	10
Total Hardness	A2340B (calculated)	1000
Water Sample - Dissolved Metals		
Aluminum	EPA 200.7	30
Sediment Sample – Metals		
Metal	**Method**	**Req. Report Limit (mg/kg - dry weight)**
Arsenic	EPA 200.8	1
Cadmium	EPA 200.8	0.2
Chromium	EPA 200.8	9
Copper	EPA 200.8	15
Iron	EPA 200.7	10
Lead	EPA 200.8	5
Zinc	EPA 200.7	20
Mercury	EPA 7471B	0.05

3.1 AQUATIC LIFE/FISH BENEFICIAL USE

Numeric standards for aquatic life support are outlined in the Circular DEQ-7 (Montana Department of Environmental Quality, 2010). Aquatic life standards are for both acute and chronic exposure. Acute criteria are based on a one hour exposure event and can only be exceeded once, on average, in a three year period. Chronic criteria are based on a 96 hour exposure and can only be exceeded, on average,

once on a three year period. There are two parts that need discussion: a) the one exceedance in three years (for both acute and chronic criteria), and b) the 96 hour exposure (for chronic criteria only).

a) One exceedance in three years: The difficulty with this approach is that attainment decisions do not take into consideration the number of results. The exceedance rate for attainment decisions will vary with the size of the data set. Thus,
   - Annual sampling- 3 samples over three years, one exceedance allowed: rate = 0.333
   - Semi-annual sampling- 6 samples over three years, one exceedance allowed: rate = 0.166
   - Quarterly sampling- 12 samples over three years, one exceedance allowed: rate = 0.0833
   - Monthly sampling- 36 samples over three years, one exceedance allowed: rate = 0.0277

   As the data set increases, the allowable exceedance rate is reduced and the potential for attainment decision errors changes. In looking at these four scenarios, the two extremes (annual sampling and monthly sampling) are not realistic due to a lack of confidence (just having 3 samples) or due to the cost (36 samples). By setting an exceedance rate of 10%, to address the "once, on average, in a three year period" (Montana Department of Environmental Quality, 2010), the level of confidence and cost would fall between the two more likely sampling scenarios of semi-annual and quarterly sampling. Multiple states have set their exceedance rate at 10% and this rate has also been recommended in past EPA guidance documents (California Environmental Protection Agency State Water Resources Control Board, 2004).

b) 96 hour exposure: the Circular DEQ-7 (Montana Department of Environmental Quality, 2010) states the chronic standard is based on a 96 hour exposure. In a lot of cases the assessor will have just one sample from a single site in any given 96 hour period, so a single sample will be used to represent the 96 hour average. In the event that there are multiple samples from the same site during a 96 hour period, then the results will need to be averaged together regardless of how many samples were collected.

For aquatic life/fish assessment, a minimum of 8 independent samples should be collected from the same assessment reach. The maximum number of samples to be collected for this assessment method is 20 samples. The maximum sample size is used as an end point so the assessor is not in an endless loop of sampling when dealing with borderline exceeding datasets. By setting a maximum sample size of 20, the alpha and beta error rate is about 35%(Drygas, Jonathan and Mark Bostrom, personal communication 2011). If the alpha and beta error rate was reduced to 25%, then 50 samples would be the maximum. If more than 20 samples have already been collected, all current data within the last ten years should be considered when applying the assessment method.

In some cases the data sets may be small and additional sampling is not feasible. In these cases, a data set of 6 samples can be used but caution should be taken when making attainment decisions with data set sizes that small. If only one sample exceeds acute or chronic standards, you will be over the ten percent threshold but the beta error rate will be very high (>65%), which means you are likely to conclude an assessment reach is in compliance when in truth it is not. A minimum of 3 exceedances will be required to determine if attainment has not been met. In this situation, if sources are identified and supporting sediment metals concentrations are exceeding the recommended sediment values, then the attainment decision will be to list or to remain listed. If no sources are identified and supporting data show exceeding the recommended sediment values (Buchman, 2008), then consult with management.

If more than 10% of the samples exceed the standard, then the attainment decision is to list or to remain listed. If the exceedance rate is equal to or less than 10%, then the attainment decision is not to list or delist. There are two exceptions to the 10% exceedance rate attainment decisions: a) if **twice** the **acute** standard is exceeded in a sample, then the attainment decision is to list or to remain listed regardless of the percent exceedance by the data set or the data set size. b) If the 10% exceedance rate threshold is surpassed but no human caused metals sources are found in the drainage, then the assessor should consult management for a case by case review.

A method for how to select independent samples and deal with larger data sets samples is in development and will be addressed at a future date.

3.2 DRINKING WATER BENEFICIAL USE (HUMAN HEALTH STANDARD)

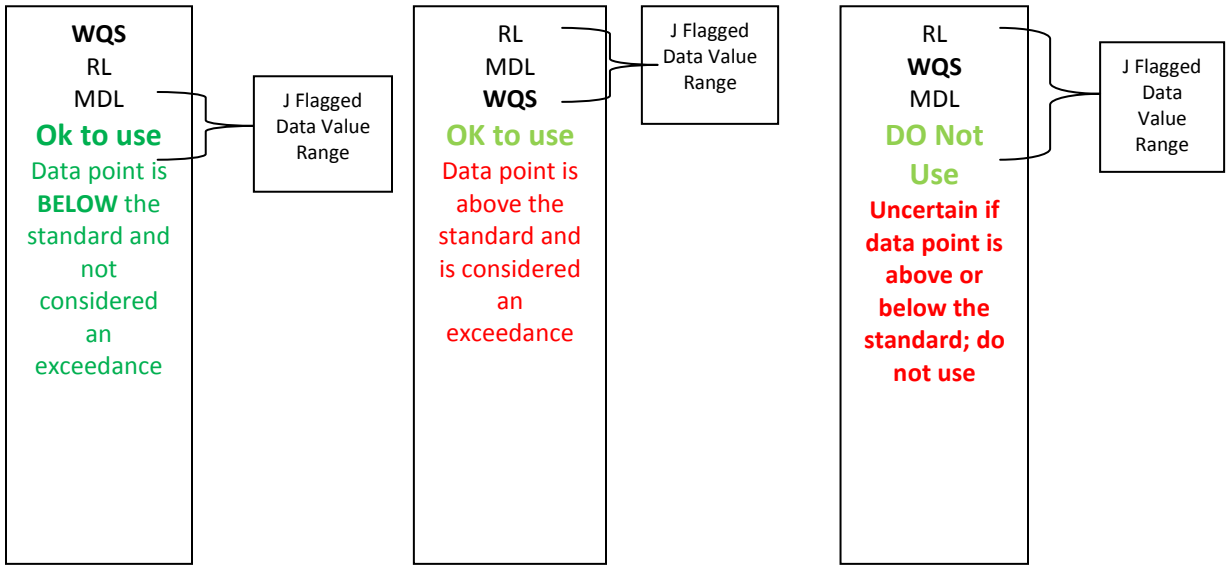
Numeric standards for human health are outlined in the Circular DEQ-7(Montana Department of Environmental Quality, 2010). No standard exceedances are allowed when assessing for human health. A minimum sample size for an assessment reach is eight samples. However, if just one sample exceeds the human health standard, then the attainment decision is to list or to remain listed. If there are no exceedances of human health standards then the attainment decision should be not to list or to delist.

3.3 NON-DETECTS AND J FLAGGED SAMPLES

The use of non-detect data points is dependent on the relation of the Water Quality Standard (WQS) to the detection limit. If the WQS is a higher value than the detection limits, then the non-detect data point can be used in the data set. If the detection limit for a non-detect data point is higher than the WQS, then it becomes uncertain if the standard is exceeded and the data point should not to be used as part of the data set.

J flagged data are to be handled in a similar manner. Data are J flagged when the result falls between the Reporting Limit (RL) and the Method Detection Limit (MDL) but the value given is considered an estimate with a low error bar. The use of J flagged data is dependent on where the WQS is in relation to the reporting limit and the method detection limit (**Figure 3-1**).

Figure 3-1. Shows three scenarios involving J flagged data and the Water Quality Standard.



The figures show that two of three possibilities result in usable data. (1) When the reporting limit is below the WQS, the J-flagged data point can be used and is not counted as an exceedance. (2) If the reporting limit and method detection limit are above the WQS, the J-flagged data point can be used and it should be counted as an exceedance. (3) Only when the standard is between the method detection limit and the reporting limit, the J-flagged data point becomes unusable. This is due to the uncertainty if the data value is above or below the standard and the value should not be used in making an impairment decision.

4.0 REFERENCES

Buchman, Michael F. 2008. NOAA Screening Quick Reference Tables. NOAA OR&R Report 08-1. Seattle, WA: Office of Repsonse and Restoration Division, national Oceanic and Atmospheric Administration.

Buck, Sharon, Walter K. Dodds, Jen Fisher, David A. Flemer, Debra Hart, Amanda K. Parker, Jan Stevenson, Vicki Watson, and Eugene B. Welch. 2000. Nutrient Criteria Technical Guidance Manual, Rivers and Streams. Washington, DC: United States Environmental Protection Agency. Report EPA-822-B00-002. http://www.epa.gov/waterscience/criteria/nutrient/guidance/rivers/index.html.

California Environmental Protection Agency State Water Resources Control Board. 2004. Functional Equivalent Document, Water Quality Control Policy for Developing California's Clean Water Act Section 303(d) List -Final. California: **California Environmental Protection Agency State Water Resources Control Board**.

Cochran, William G. 1977. Sampling Techniques, 3 ed., New York: John Wiley and Sons.

Drygas, Jonathan and Mark Bostrom. 2011. Personal Communication.

Environmental Research Laboratory-Duluth. 2002. Consolidated Assessment and Listing Methodology: Towards a Compendium of Best Practices. Washington, D.C.: U.S. Envrionmental Protection Agency.

Flynn, Kyle and Michael W. Suplee. 2010. Defining Large Rivers in Montana Using a Wadeability Index. Helena, MT: Montana Department of Environmental Quality. http://deq.mt.gov/wqinfo/standards/default.mcpx. Accessed 6/9/11.

Hurlbert, Stuart H. 1984. Pseudoreplication and the Design of Ecological Field Experiments. *Ecological Monographs.* 54(2): 187-211.

Montana Department of Environmental Quality. 2005. Field Procedures Manual For Water Quality Assessment Monitoring. Helena, MT: Montana Department of Environmental Quality, Water Quality Planning Bureau. Report WQPBWQM-020.

-----. 2010. Circular DEQ-7: Montana Numeric Water Quality Standards. Helena, MT: Montana Department of Environmental Quality. http://deq.mt.gov/wqinfo/Standards/PDF/DEQ-7.pdf. Accessed 6/9/11.

Norris, Richard H., E. P. McElravy, and V. H. Resh. 1992. "The River Handbook,", Calow, Peter and Petts, Jeffrey E., Ch. The Sampling Problem, (Oxford, England: Blackwell Scientific Publications)

Suplee, Michael W. and Rosie Sada de Suplee. 2011. Assessment Methodology for Determing Wadeable Stream Impairment Due to Excess Nitrogen and Phosphorus Levels. Helena, MT: Montana Department of Environmental Quality.

U.S. Environmental Protection Agency. 2002. Consolidated Assessment and Listing Methodology: Towards a Compendium of Best Practices. Washington, D.C.: U.S. Envrionmental Protection Agency.